

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/530,556  
Source: PLT  
Date Processed by STIC: 05/12/2006

# ***ENTERED***



PCT

## RAW SEQUENCE LISTING

DATE: 05/12/2006

PATENT APPLICATION: US/10/530,556

TIME: 08:55:29

Input Set : A:\781\_2\_PCT\_SEQLIST.TXT

Output Set: N:\CRF4\05122006\J530556.raw

```

4 <110> APPLICANT: Genencor International, Inc.
5     Dunn-Coleman, Nigel
6     Ward, Michael
8 <120> TITLE OF INVENTION: BGL6 Beta-Glucosidase and Nucleic Acids
9     Encoding the Same
11 <130> FILE REFERENCE: GC781-2-PCT
C--> 13 <140> CURRENT APPLICATION NUMBER: US/10/530,556
C--> 14 <141> CURRENT FILING DATE: 2005-04-07
16 <150> PRIOR APPLICATION NUMBER: US 60/424,784
17 <151> PRIOR FILING DATE: 2002-11-07
19 <160> NUMBER OF SEQ ID NOS: 4
21 <170> SOFTWARE: FastSEQ for Windows Version 4.0
23 <210> SEQ ID NO: 1
24 <211> LENGTH: 2812
25 <212> TYPE: DNA
26 <213> ORGANISM: Trichoderma reesei
28 <400> SEQUENCE: 1
29 gatcacacccc ctccccaccct tctctttttca aggttgtccc cttctcccac ggctttatgt      60
30 acttccccact ctmtaattcg ctctttccat tccaagccaa gcaacatctg tgagcagctc      120
31 atccttccca atattggcga atggcaggag cagatgatgg gttttgacgt ggaggatggt      180
32 ctgtctcagc tgagccaaaa tgagaagatt gctctcttgt ccggcattga tttctggcat      240
33 acttatccca taccaaagta caacgtccct tcagtccgcc taacggacgg tcctaacggc      300
34 atacgaggca caaagttttt tgctggcatt cctgctgcct gcctgccatg tgggacggcc      360
35 ctggcctcta cctgggataa gcagctgctg aagaaggctg ggaagctgct cggatgatgag      420
36 tgcatcgcaa aaggcgccca ctgctggctg ggcccaacaa tcaatactcc ccgatctcct      480
37 ctggggggggc gcggcttcga gtcattttcg gaagatccgt acctgtccgg catccttgct      540
38 gcatctatga ttctcggtcg tgaaagcaca ggtgtcatct ctgccgtcaa acactttgtc      600
39 gccaacgacc aggagcacga gcggcgagcg gtcgactgtc tcatcaccca gcgggctctc      660
40 cgggaggtct atctgcgacc cttccagatc gtagcccgag atgcaaggcc cggcgcatgt      720
41 atgacatcct acaacaaggt caatggcaag cacgtcgctg acagcgccga gttccttcag      780
42 ggcattctcc ggactgagtg gaattgggat cctctcattg tcagcgactg gtacggcacc      840
43 tacaccacta ttgatgccat caaagccggc cttgatctcg agatgccggg cgtttcacga      900
44 tatcgcgga aatacatcga gtctgctctg caggcccggt tgctgaagca gtccactatc      960
45 gatgagcgcg ctgcgcgcgt gctcaggttc gccacagaag ccagccatct caaggctctc      1020
46 gaggtagagc aaggcgcgtga cttcccagag gatcgcgctc tcaaccgtca gatctgcggc      1080
47 agcagcattg tcctactgaa gaatgagaac tccatcttac ctctcccca gtcggtcaag      1140
48 aaggctcgcc ttgttggatc ccacgtgcgt ctaccggcta tctcgggagg aggcagcgcc      1200
49 tctcttgctc cttactatgc catatctcta tacgatgccg tctctgaggt actagccggg      1260
50 gccacgatca cgcacgaggt cgggtgcctat gccacacaaa tgctgcccgt catcgacgca      1320
51 atgatcagca acgcccgaat ccacttctac aacgacccca tcgatgtcaa agacagaaag      1380
52 ctccctggca gtgagaacgt atcgctcgaca tcgttcagc tcatggatta caacaacatc      1440
53 ccaacgctca acaaggccat gttctggggg actctcgtgg gcgagtttat ccctaccgcc      1500
54 acgggaattt ggggaatttg cctcagtgctc tttggcactg ccgacctta tattgataat      1560

```

## RAW SEQUENCE LISTING

DATE: 05/12/2006

PATENT APPLICATION: US/10/530,556

TIME: 08:55:29

Input Set : A:\781\_2\_PCT\_SEQLIST.TXT

Output Set: N:\CRF4\05122006\J530556.raw

```

55 gagctcgtga ttgaaaatac aacacatcag acgcgtggta ccgccttttt cggaaaggga 1620
56 acgacggaaa aagtcgctac caggaggatg gtggccggca gcacctacaa gctgcgtctc 1680
57 gagtttgggt ctgccaacac gaccaagatg gagacgaccg gtgttgtaa ctttggcggc 1740
58 ggtgccgtac acctgggtgc ctgtctcaag gtcgaccac aggagatgat tgcgcgggcc 1800
59 gtcaaggccg cagccgatgc cgactacacc atcatctgca cgggactcag cggcgagtgg 1860
60 gagtctgagg gttttgaccg gcctcacatg gacctgcccc ctggtgtgga caccatgatc 1920
61 tcgcaagttc ttgacgccgc tcccaatgct gtagtcgtca accagtcagg caccctcagt 1980
62 acaatgagct gggctcataa agcaaaggcc attgtgcagg ctggttatgg tggtaacgag 2040
63 acaggccacg gaatctccga tgtgctcttt ggcaacgtca acccgtcggg gaaactctcc 2100
64 ctatcgtggc cagtcgatgt gaagcacaac ccagcatatc tcaactacgc cagcgttggt 2160
65 ggacgggtct tgtatggcga ggatgtttac gttggctaca agttctacga caaaacggag 2220
66 agggagggtc tgtttccttt tgggcatggc ctgtcttacg ctaccttcaa gctcccagat 2280
67 tctaccgtga ggacggtccc cgaaaccttc caccggacc agcccacagt agccattgtc 2340
68 aagatcaaga acacgagcag tgtcccgggc gccagggtcc tgcagctata catttcggcc 2400
69 ccaaactcgc ctacacatcg cccggtcaag gagctgcacg gattcgaaaa ggtgtatctt 2460
70 gaagctggcg aggagaagga ggtacaaata cccattgacc agtacgctac tagcttctgg 2520
71 gacgagattg agagcatgtg gaagagcgag aggggcattt atgatgtgct tgtaggattc 2580
72 tcgagtcagg aaatctcggg caaggggaag ctgattgtgc ctgaaacgcg attctggatg 2640
73 gggctgtaga ttcaacacgt gagcaaaagc gattgctgaa agtaccagaa aagccaaggg 2700
74 agtcaaagga tgggaacttg tgtcaataga agatatgcat gatgggcatt tgggatggg 2760
75 tttggcatta tgcaaagaag caaagatgga gtgataaaaa aaaaaaaaaa aa 2812

```

77 &lt;210&gt; SEQ ID NO: 2

78 &lt;211&gt; LENGTH: 838

79 &lt;212&gt; TYPE: PRT

80 &lt;213&gt; ORGANISM: Trichoderma reesei

82 &lt;400&gt; SEQUENCE: 2

```

83 Met Gly Glu Trp Gln Glu Gln Met Met Gly Phe Asp Val Glu Asp Val
84 1 5 10 15
85 Leu Ser Gln Leu Ser Gln Asn Glu Lys Ile Ala Leu Leu Ser Gly Ile
86 20 25 30
87 Asp Phe Trp His Thr Tyr Pro Ile Pro Lys Tyr Asn Val Pro Ser Val
88 35 40 45
89 Arg Leu Thr Asp Gly Pro Asn Gly Ile Arg Gly Thr Lys Phe Phe Ala
90 50 55 60
91 Gly Ile Pro Ala Ala Cys Leu Pro Cys Gly Thr Ala Leu Ala Ser Thr
92 65 70 75 80
93 Trp Asp Lys Gln Leu Leu Lys Lys Ala Gly Lys Leu Leu Gly Asp Glu
94 85 90 95
95 Cys Ile Ala Lys Gly Ala His Cys Trp Leu Gly Pro Thr Ile Asn Thr
96 100 105 110
97 Pro Arg Ser Pro Leu Gly Gly Arg Gly Phe Glu Ser Phe Ser Glu Asp
98 115 120 125
99 Pro Tyr Leu Ser Gly Ile Leu Ala Ala Ser Met Ile Leu Gly Cys Glu
100 130 135 140
101 Ser Thr Gly Val Ile Ser Ala Val Lys His Phe Val Ala Asn Asp Gln
102 145 150 155 160
103 Glu His Glu Arg Arg Ala Val Asp Cys Leu Ile Thr Gln Arg Ala Leu
104 165 170 175
105 Arg Glu Val Tyr Leu Arg Pro Phe Gln Ile Val Ala Arg Asp Ala Arg

```

## RAW SEQUENCE LISTING

DATE: 05/12/2006

PATENT APPLICATION: US/10/530,556

TIME: 08:55:29

Input Set : A:\781\_2\_PCT\_SEQLIST.TXT

Output Set: N:\CRF4\05122006\J530556.raw

```

106          180          185          190
107 Pro Gly Ala Leu Met Thr Ser Tyr Asn Lys Val Asn Gly Lys His Val
108          195          200          205
109 Ala Asp Ser Ala Glu Phe Leu Gln Gly Ile Leu Arg Thr Glu Trp Asn
110          210          215          220
111 Trp Asp Pro Leu Ile Val Ser Asp Trp Tyr Gly Thr Tyr Thr Thr Ile
112 225          230          235          240
113 Asp Ala Ile Lys Ala Gly Leu Asp Leu Glu Met Pro Gly Val Ser Arg
114          245          250          255
115 Tyr Arg Gly Lys Tyr Ile Glu Ser Ala Leu Gln Ala Arg Leu Leu Lys
116          260          265          270
117 Gln Ser Thr Ile Asp Glu Arg Ala Arg Arg Val Leu Arg Phe Ala Gln
118          275          280          285
119 Lys Ala Ser His Leu Lys Val Ser Glu Val Glu Gln Gly Arg Asp Phe
120          290          295          300
121 Pro Glu Asp Arg Val Leu Asn Arg Gln Ile Cys Gly Ser Ser Ile Val
122 305          310          315          320
123 Leu Leu Lys Asn Glu Asn Ser Ile Leu Pro Leu Pro Lys Ser Val Lys
124          325          330          335
125 Lys Val Ala Leu Val Gly Ser His Val Arg Leu Pro Ala Ile Ser Gly
126          340          345          350
127 Gly Gly Ser Ala Ser Leu Val Pro Tyr Tyr Ala Ile Ser Leu Tyr Asp
128          355          360          365
129 Ala Val Ser Glu Val Leu Ala Gly Ala Thr Ile Thr His Glu Val Gly
130          370          375          380
131 Ala Tyr Ala His Gln Met Leu Pro Val Ile Asp Ala Met Ile Ser Asn
132 385          390          395          400
133 Ala Val Ile His Phe Tyr Asn Asp Pro Ile Asp Val Lys Asp Arg Lys
134          405          410          415
135 Leu Leu Gly Ser Glu Asn Val Ser Ser Thr Ser Phe Gln Leu Met Asp
136          420          425          430
137 Tyr Asn Asn Ile Pro Thr Leu Asn Lys Ala Met Phe Trp Gly Thr Leu
138          435          440          445
139 Val Gly Glu Phe Ile Pro Thr Ala Thr Gly Ile Trp Glu Phe Gly Leu
140          450          455          460
141 Ser Val Phe Gly Thr Ala Asp Leu Tyr Ile Asp Asn Glu Leu Val Ile
142 465          470          475          480
143 Glu Asn Thr Thr His Gln Thr Arg Gly Thr Ala Phe Phe Gly Lys Gly
144          485          490          495
145 Thr Thr Glu Lys Val Ala Thr Arg Arg Met Val Ala Gly Ser Thr Tyr
146          500          505          510
147 Lys Leu Arg Leu Glu Phe Gly Ser Ala Asn Thr Thr Lys Met Glu Thr
148          515          520          525
149 Thr Gly Val Val Asn Phe Gly Gly Gly Ala Val His Leu Gly Ala Cys
150          530          535          540
151 Leu Lys Val Asp Pro Gln Glu Met Ile Ala Arg Ala Val Lys Ala Ala
152 545          550          555          560
153 Ala Asp Ala Asp Tyr Thr Ile Ile Cys Thr Gly Leu Ser Gly Glu Trp
154          565          570          575

```

## RAW SEQUENCE LISTING

DATE: 05/12/2006

PATENT APPLICATION: US/10/530,556

TIME: 08:55:29

Input Set : A:\781\_2\_PCT\_SEQLIST.TXT

Output Set: N:\CRF4\05122006\J530556.raw

```

155 Glu Ser Glu Gly Phe Asp Arg Pro His Met Asp Leu Pro Pro Gly Val
156          580          585          590
157 Asp Thr Met Ile Ser Gln Val Leu Asp Ala Ala Pro Asn Ala Val Val
158          595          600          605
159 Val Asn Gln Ser Gly Thr Pro Val Thr Met Ser Trp Ala His Lys Ala
160          610          615          620
161 Lys Ala Ile Val Gln Ala Trp Tyr Gly Gly Asn Glu Thr Gly His Gly
162 625          630          635          640
163 Ile Ser Asp Val Leu Phe Gly Asn Val Asn Pro Ser Gly Lys Leu Ser
164          645          650          655
165 Leu Ser Trp Pro Val Asp Val Lys His Asn Pro Ala Tyr Leu Asn Tyr
166          660          665          670
167 Ala Ser Val Gly Gly Arg Val Leu Tyr Gly Glu Asp Val Tyr Val Gly
168          675          680          685
169 Tyr Lys Phe Tyr Asp Lys Thr Glu Arg Glu Val Leu Phe Pro Phe Gly
170          690          695          700
171 His Gly Leu Ser Tyr Ala Thr Phe Lys Leu Pro Asp Ser Thr Val Arg
172 705          710          715          720
173 Thr Val Pro Glu Thr Phe His Pro Asp Gln Pro Thr Val Ala Ile Val
174          725          730          735
175 Lys Ile Lys Asn Thr Ser Ser Val Pro Gly Ala Gln Val Leu Gln Leu
176          740          745          750
177 Tyr Ile Ser Ala Pro Asn Ser Pro Thr His Arg Pro Val Lys Glu Leu
178          755          760          765
179 His Gly Phe Glu Lys Val Tyr Leu Glu Ala Gly Glu Glu Lys Glu Val
180          770          775          780
181 Gln Ile Pro Ile Asp Gln Tyr Ala Thr Ser Phe Trp Asp Glu Ile Glu
182 785          790          795          800
183 Ser Met Trp Lys Ser Glu Arg Gly Ile Tyr Asp Val Leu Val Gly Phe
184          805          810          815
185 Ser Ser Gln Glu Ile Ser Gly Lys Gly Lys Leu Ile Val Pro Glu Thr
186          820          825          830
187 Arg Phe Trp Met Gly Leu
188          835
190 <210> SEQ ID NO: 3
191 <211> LENGTH: 2517
192 <212> TYPE: DNA
193 <213> ORGANISM: Trichoderma reesei
195 <400> SEQUENCE: 3
196 atgggcgaat ggcaggagca gatgatgggt tttgacgtgg aggatgttct gtctcagctg      60
197 agccaaaatg agaagattgc tctcttgtcc ggcattgatt tctggcatac ttatcccata      120
198 ccaaagtaca acgtcccttc agtccgccta acggacgggtc ctaacggcat acgaggcaca      180
199 aagttttttg ctggcattcc tgetgcctgc ctgccatgtg ggacggccct ggccctctacc      240
200 tgggataagc agctgctgaa gaaggctggg aagctgctcg gtgatgagtg catcgcaaaa      300
201 ggcgcccact gctggctggg cccaacaatc aatactcccc gatctcctct gggggggcgc      360
202 ggcttcgagt cattttcgga agatccgtac ctgtccggca tccttgctgc atctatgatt      420
203 ctcggtgtg aaagcacagg tgtcatctct gccgtcaaac actttgtcgc caacgaccag      480
204 gagcacgagc ggcgagcggc cgactgtctc atcaccacgc gggctctccg ggaggtctat      540
205 ctgcgaccct tccagatcgt agcccagat gcaaggcccc gcgcattgat gacatcctac      600

```

## RAW SEQUENCE LISTING

DATE: 05/12/2006

PATENT APPLICATION: US/10/530,556

TIME: 08:55:29

Input Set : A:\781\_2\_PCT\_SEQLIST.TXT

Output Set: N:\CRF4\05122006\J530556.raw

```

206 aacaagggtca atggcaagca cgtcgtgac agcgccgagt tccttcaggg cattctccgg 660
207 actgagtgga attgggatcc tctcattgtc agcgactggt acggcaccta caccactatt 720
208 gatgccatca aagccggcct tgatctcgag atgccgggagc ttctacgata tcgcggaaca 780
209 tacatcgagt ctgctctgca ggcccgcttg ctgaagcagt ccactatcga tgagcgcgct 840
210 cgccgctgac tcagggttcgc ccagaaggcc agccatctca aggtctccga ggtagagcaa 900
211 ggccgtgact tcccagagga tcgctgcttc aaccgtcaga tctgcggcag cagcattgtc 960
212 ctactgaaga atgagaactc catcttacct ctccccaagt ccgtcaagaa ggtcgccctt 1020
213 gttggatccc acgtgcgtct accggctatc tcgggaggag gcagcgctc tcttgccct 1080
214 tactatgcca tatctctata cgatgccgtc tctgaggtag tagccggtgc cagcatcacg 1140
215 cagcagggtcg gtgcctatgc ccaccaaag ctgcccgtca tcgacgcaat gatcagcaac 1200
216 gccgtaatcc acttctacaa cgaccccatc gatgtcaaag acagaaagct ccttggcagt 1260
217 gagaacgtat cgtcgacatc gttccagctc atggattaca acaacatccc aacgctcaac 1320
218 aaggccatgt tctggggtag tctcgtgggc gagtttatcc ctaccgccac gggaatttgg 1380
219 gaatttggcc tcagtgtctt tggcactgcc gacctttata ttgataatga gctcgtgatt 1440
220 gaaaatacaa cacatcagac gcgtggtacc gcctttttcg gaaagggaaac gacggaaaaa 1500
221 gtcgctacca ggaggatggt ggccggcagc acctacaagc tgcgtctcga gtttgggtct 1560
222 gccaacacga ccaagatgga gacgaccggt gttgtcaact ttggcggcgg tgccgtacac 1620
223 ctgggtgcct gtctcaaggc cgaccacacg gagatgattg cgcgggcccgt caaggccgca 1680
224 gccgatgccg actacaccat catctgcacg ggactcagcg gcgagtggga gtctgagggt 1740
225 ttgaccggc ctcacatgga cctgccccct ggtgtggaca ccatgatctc gcaagttctt 1800
226 gacgccgctc ccaatgctgt agtcgtcaac cagtcaggca cccagtgac aatgagctgg 1860
227 gctcataaag caaaggccat tgtgcaggct tggtaggtg gtaacgagac agggccacgga 1920
228 atctccgatg tgctcttttg caacgtcaac ccgtcgggga aactctccct atcgtggcca 1980
229 gtcgatgtga agcacaacc agcatatctc aactacgcca gcgttggtgg acgggtcttg 2040
230 tatggcgagg atgtttacgt tggctacaag ttctacgaca aaacggagag ggaggttctg 2100
231 ttctcttttg ggcattggct gtcttacgct accttcaagc tcccagattc taccgtgagg 2160
232 acggtccccg aaaccttcca cccggaccag cccacagtag ccattgtcaa gatcaagaac 2220
233 acgagcagtg tcccgggcgc ccaggctctg cagctataca ttccggcccc aaactcgct 2280
234 acacatcgcc cgggtcaagga gctgcacgga ttcgaaaagg tgtatcttga agctggcgag 2340
235 gagaaggagg tacaaatacc cattgaccag tacgctacta gcttctggga cgagattgag 2400
236 agcatgtgga agagcgagag gggcatttat gatgtgcttg taggattctc gagtccaggaa 2460
237 atctcgggca aggggaagct gattgtgcct gaaacgcgat tctggatggg gctgtag 2517
239 <210> SEQ ID NO: 4
240 <211> LENGTH: 831
241 <212> TYPE: PRT
242 <213> ORGANISM: Trichoderma reesei
244 <400> SEQUENCE: 4
245 Met Met Gly Phe Asp Val Glu Asp Val Leu Ser Gln Leu Ser Gln Asn
246 1 5 10 15
247 Glu Lys Ile Ala Leu Leu Ser Gly Ile Asp Phe Trp His Thr Tyr Pro
248 20 25 30
249 Ile Pro Lys Tyr Asn Val Pro Ser Val Arg Leu Thr Asp Gly Pro Asn
250 35 40 45
251 Gly Ile Arg Gly Thr Lys Phe Ala Gly Ile Pro Ala Ala Cys Leu
252 50 55 60
253 Pro Cys Gly Thr Ala Leu Ala Ser Thr Trp Asp Lys Gln Leu Leu Lys
254 65 70 75 80
255 Lys Ala Gly Lys Leu Leu Gly Asp Glu Cys Ile Ala Lys Gly Ala His
256 85 90 95

```

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/530,556

DATE: 05/12/2006

TIME: 08:55:30

Input Set : A:\781\_2\_PCT\_SEQLIST.TXT

Output Set: N:\CRF4\05122006\J530556.raw

L:13 M:270 C: Current Application Number differs, Replaced Current Application Number

L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date